

VOICE CONTROLLED COMPUTER INTERFACE

Background of the Invention

This invention relates to voice controlled computer interfaces.

5 Voice recognition systems can convert human speech into computer information. Such voice recognition systems have been used, for example, to control text-type user interfaces, e.g., the text-type interface of the disk operating system (DOS) of the IBM Personal Computer.

10 Voice control has also been applied to graphical user interfaces, such as the one implemented by the Apple Macintosh computer, which includes icons, pop-up windows, and a mouse. These voice control systems use voiced commands to generate keyboard keystrokes.

Summary of the Invention

15 In general, in one aspect, the invention features enabling voiced utterances to be substituted for manipulation of a pointing device, the pointing device being of the kind which is manipulated to control motion of a cursor on a computer display and to indicate desired actions associated with the position of the cursor on the display, the cursor being moved and the desired actions being aided by an operating system in the computer in response to control signals received from the pointing device, the computer also having an alphanumeric keyboard, the operating system being separately
20 responsive to control signals received from the keyboard in accordance with a predetermined format specific to the keyboard; a voice recognizer recognizes the voiced utterance, and an interpreter converts the voiced utterance into control signals which will directly create a desired action aided by the operating
25 system without first being converted into control signals expressed in the predetermined format specific to the keyboard.

30 In general, in another aspect of the invention, voiced utterances are converted to commands, expressed in a predefined command language, to be used by an operating system of a computer,
35 converting some voiced utterances into commands corresponding to

actions to be taken by said operating system, and converting other voiced utterances into commands which carry associated text strings to be used as part of text being processed in an application program running under the operating system.

5 In general, in another aspect, the invention features generating a table for aiding the conversion of voiced utterances to commands for use in controlling an operating system of a computer to achieve desired actions in an application program running under the operating system, the application program including menus and control buttons; the instruction sequence of the application program is parsed to identify menu entries and control buttons, and an entry is included in the table for each menu entry and control button found in the application program, each entry in the table containing a command corresponding to the menu entry or control button.

10 In general, in another aspect, the invention features enabling a user to create an instance in a formal language of the kind which has a strictly defined syntax; a graphically displayed list of entries are expressed in a natural language and do not comply with the syntax, the user is permitted to point to an entry on the list, and the instance corresponding to the identified entry in the list is automatically generated in response to the pointing.

20 The invention enables a user to easily control the graphical interface of a computer. Any actions that the operating system can be commanded to take can be commanded by voiced utterances. The commands may include commands that are normally entered through the keyboard as well as commands normally entered through a mouse or any other input device. The user may switch back and forth between voiced utterances that correspond to commands for actions to be taken and voiced utterances that correspond to text strings to be used in an application program without giving any indication that the switch has been made. Any application may be made susceptible to a voice interface by automatically parsing the application instruction sequence for menus and control buttons that control the application.

Other advantages and features will become apparent from the following description of the preferred embodiment and from the claims.

Description of the Preferred Embodiment

We first briefly describe the drawings.

Fig. 1 is a functional block diagram of a Macintosh computer served by a Voice Navigator voice controlled interface system.

Fig. 2A is a functional block diagram of a Language Maker system for creating word lists for use with the Voice Navigator interface of Fig. 1.

Fig. 2B depicts the format of the voice files and word lists used with the Voice Navigator interface.

Fig. 3 is an organizational block diagram of the Voice Navigator interface system.

Fig. 4 is a flow diagram of the Language Maker main event loop.

Fig. 5 is a flow diagram of the Run Edit module.

Fig. 6 is a flow diagram of the Record Actions submodule.

Fig. 7 is a flow diagram of the Run Modal module.

Fig. 8 is a flow diagram of the In Button? routine.

Fig. 9 is a flow diagram of the Event Handler module.

Fig. 10 is a flow diagram of the Do My Menu module.

Figs. 11A through 11I are flow diagrams of the Language Maker menu submodules.

Fig. 12 is a flow diagram of the Write Production module.

Fig. 13 is a flow diagram of the Write Terminal submodule.

Fig. 14 is a flow diagram of the Voice Control main driver loop.

Fig. 15 is a flow diagram of the Process Input module.

Fig. 16 is a flow diagram of the Recognize submodule.

Fig. 17 is a flow diagram of the Process Voice Control Commands routine.

Fig. 18 is a flow diagram of the ProcessQ module.

Fig. 19 is a flow diagram of the Get Next submodule.

Fig. 20 is a chart of the command handlers.

Figs. 21A through 21G are flow diagrams of the command handlers.

Fig. 22 is a flow diagram of the Post Mouse routine.

Fig. 23 is a flow diagram of the Set Mouse Down routine.

5 Figs. 24 and 25 illustrate the screen displays of Voice Control.

Figs. 26 through 29 illustrate the screen displays of Language Maker.

Fig. 30 is a listing of a language file.

10 System Overview

Referring to Fig. 1, in an Apple Macintosh computer 100, a Macintosh operating system 132 provides a graphical interactive user interface by processing events received from a mouse 134 and a keyboard 136 and by providing displays including icons, windows, and menus on a display device 138. Operating system 132 provides an environment in which application programs such as MacWrite 139, desktop utilities such as Calculator 137, and a wide variety of other programs can be run.

20 The operating system 132 also receives events from the Voice Navigator voice controlled computer interface 102 to enable the user to control the computer by voiced utterances. For this purpose, the user speaks into a microphone 114 connected via a Voice Navigator box 112 to the SCSI (Small Computer Systems Interface) port of the computer 100. The Voice Navigator box 112 digitizes and processes analog audio signals received from a microphone 114, and transmits processed digitized audio signals to the Macintosh SCSI port. The Voice Navigator box includes an analog-to-digital converter (A/D) for digitizing the audio signal, a DSP (Digital Signal Processing) chip for compressing the resulting digital samples, and protocol interface hardware which configures the digital samples to obey the SCSI protocols.

30 Recognizer Software 120 (available from Dragon Systems, Newton, MA) runs under the Macintosh operating system, and is controlled by internal commands 123 received from Voice Control driver 128 (which also operates under the Macintosh operating

35

system). One possible algorithm for implementing Recognizer Software 120 is disclosed by Baker et al, in US Patent 4,783,803, incorporated by reference herein. Recognizer Software 120 processes the incoming compressed, digitized audio, and compares each utterance of the user to prestored utterance macros. If the user utterance matches a prestored utterance macro, the utterance is recognized, and a command string 121 corresponding to the recognized utterance is delivered to a text buffer 126. Command strings 121 delivered from the Recognizer Software represent commands to be issued to the Macintosh operating system (e.g., menu selections to be made or text to be displayed), or internal commands 123 to be issued by the Voice Control driver.

During recognition, the Recognizer Software 120 compares the incoming samples of an utterance with macros in a voice file 122. (The system requires the user to space apart his utterances briefly so that the system can recognize when each utterance ends.) The voice file macros are created by a "training" process, described below. If a match is found (as judged by the recognition algorithm of the Recognizer Software 120), a Voice Control command string from a word list 124 (which has been directly associated with voice file 122) is fetched and sent to text buffer 126.

The command strings in text buffer 126 are relayed to Voice Control driver 128, which drives a Voice Control interpreter 130 in response to the strings.

A command string 121 may indicate an internal command 123, such as a command to the Recognizer Software to "learn" new voice file macros, or to adjust the sensitivity of the recognition algorithm. In this case, Voice Control interpreter 130 sends the appropriate internal command 123 to the Recognizer Software 120. In other cases, the command string may represent an operating system manipulation, such as a mouse movement. In this case, Voice Control interpreter 130 produces the appropriate action by interacting with the Macintosh operating system 132.

Each application or desktop accessory is associated with a word list 124 and a corresponding voice file 122; these are loaded

by the Recognition Software when the application or desktop accessory is opened.

The voice files are generated by the Recognizer Software 120 in its "learn" mode, under the control of internal commands from the Voice Control driver 128.

The word lists are generated by the Language Maker desktop accessory 140, which creates "languages" of utterance names and associated Voice Control command strings, and converts the languages into the word lists. Voice Control command strings are strings such as "ESC", "TEXT", "@MENU(font,2)", and belong to a Voice Control command set, the syntax of which will be described later and is set forth in Appendix A.

The Voice Control and Language Maker software includes about 30,000 lines of code, most of which is written in the C language, the remainder being written in assembly language. A listing of the Voice Control and Language Maker software is provided in microfiche as appendix C. The Voice Control software will operate on a Macintosh Plus or later models, configured with a minimum of 1 Mbyte RAM (2 Mbyte for HyperCard and other large applications), a Hard Disk, and with Macintosh operating system version 6.01 or later.

In order to understand the interaction of the Voice Control interpreter 130 and the operating system, note that Macintosh operating system 132 is "event driven". The operating system maintains an event queue (not shown); input devices such as the mouse 134 or the keyboard 136 "post" events to this queue to cause the operating system to, for example, create the appropriate text entry, or trigger a mouse movement. The operating system 132 then, for example, passes messages to Macintosh applications (such as MacWrite 139) or to desktop accessories (such as Calculator 137) indicating events on the queues (if any). In one mode of operation, Voice Control interpreter 130 likewise controls the operating system (and hence the applications and desktop accessories which are currently running) by posting events to the operating system queues. The events posted by the Voice Control

interpreter typically correspond to mouse activity or to keyboard keystrokes, or both, depending upon the voice commands. Thus, the Voice Navigator system 102 provides an additional user interface. In some cases, the "voice" events may comprise text strings to be displayed or included with text being processed by the application program.

At any time during the operation of the Voice Navigator system, the Recognizer Software 120 may be trained to recognize an utterance of a particular user and to associate a corresponding text string with each utterance. In this mode, the Recognizer Software 120 displays to the user a menu of the utterance names (such as "file", "page down") which are to be recognized. These names, and the corresponding Voice Control command strings (indicating the appropriate actions) appear in a current word list 124. The user designates the utterance name of interest and then is prompted to speak the utterance corresponding to that name. For example, if the utterance name is "file", the user might utter "FILE" or "PLEASE FILE". The digitized samples from the Voice Navigator box 112 corresponding to that utterance are then used by the Recognizer Software 120 to create a "macro" representing the utterance, which is stored in the voice file 122 and subsequently associated with the utterance name in the word list 124. Ordinarily, the utterance is repeated more than once, in order to create a macro for the utterance that accommodates variation in a particular speaker's voice.

The meaning of the spoken utterance need not correspond to the utterance name, and the text of the utterance name need not correspond to the Voice Control command strings stored in the word list. For example, the user may wish a command string that causes the operating system to save a file to have the utterance name "save file"; the associated command string may be "@MENU(file,2)"; and the utterance that the user trains for this utterance name may be the spoken phrase "immortalize". The Recognizer Software and Voice Control cause that utterance, name, and command string to be properly associated in the voice file and word list 124.

Referring to Fig. 2A, the word lists 124 used by the Voice Navigator are created by the Language Maker desk accessory 140 running under the operating system. Each word list 124 is hierarchical, that is, some utterance names in the list link to sub-lists of other utterance names. Only the list of utterance names at a currently active level of the hierarchy can be recognized. (In the current embodiment, the number of utterance names at each level of the hierarchy can be as large as 1000.) In the operation of Voice Control, some utterances, such as "file", may summon the file menu on the screen, and link to a subsequent list of utterance names at a lower hierarchical level. For example, the file menu may list subsequent commands such as "save", "open", or "save as", each associated with an utterance.

Language Maker enables the user to create a hierarchical language of utterance names and associated command strings, rearrange the hierarchy of the language, and add new utterance names. Then, when the language is in the form that the user desires, the language is converted to a word list 124. Because the hierarchy of the utterance names and command strings can be adjusted, when using the Voice Navigator system the user is not bound by the preset menu hierarchy of an application. For example, the user may want to create a "save" command at the top level of the utterance hierarchy that directly saves a file without first summoning the file menu. Also, the user may, for example, create a new utterance name "goodbye", that saves a file and exits all at once.

Each language created by Language Maker 140 also contains the command strings which represent the actions (e.g. clicking the mouse at a location, typing text on the screen) to be associated with utterances and utterance names. In order for the training of the Voice Navigator system to be more intuitive, the user does not specify the command strings to describe the actions he wishes to be associated with an utterance and utterance name. In fact, the user does not need to know about, and never sees, the command strings stored in the Language Maker language or the resulting word

list 124.

In a "record" mode, to associate a series of actions with an utterance name, the user simply performs the desired actions (such as typing the text at the keyboard, or clicking the mouse at a menu). The actions performed are converted into the appropriate command strings, and when the user turns off the record mode, the command strings are associated with the selected utterance name.

While using Language Maker, the user can cause the creation of a language by entering utterance names by typing the names at the keyboard 142, by using a "create default text" procedure 146 (to parse a text file on the clipboard, in which case one utterance name is created for each word in the text file, and the names all start at the same hierarchical level), or by using a "create default menus" procedure (to parse the executable code 144 for an application, and create a set of utterance names which equal the names of the commands in the menus of the application, in which case the initial hierarchy for the names is the same as the hierarchy of the menus in the application).

If the names are typed at the keyboard or created by parsing a text file, the names are initially associated with the keystrokes which, when typed at the keyboard, produce the name. Therefore, the name "text" would be initially be associated with the keystrokes t-e-x-t. If the names are created by parsing the executable code 144 for an application, then the names are initially associated with the command strings which execute the corresponding menu commands for the application. These initial command strings can be changed by simply selecting the utterance name to be changed and putting Language Maker into record mode.

The output of Language Maker is a language file 148. This file contains the utterance names and the corresponding command strings. The language file 148 is formatted for input to a VOCAL compiler 150 (available from Dragon Systems), which converts the language file into a word list 124 for use with the Recognition Software. The syntax of language files is specified in the Voice Navigator Developer's Reference Manual, provided as Appendix D, and

incorporated by reference.

Referring to Fig. 2B, a macro 147 of each learned utterance is stored in the voice file 122. A corresponding utterance name 149 and command string 151 are associated with one another and with the utterance and are stored in the word list 124. The word list 124 is created and modified by Language Maker 140, and the voice file 122 is created and modified by the Recognition Software 120 in its learn mode, under the control of the Voice Control driver 128.

Referring to Fig. 3, in the Voice Navigator system 102, the Voice Navigator hardware box 152 includes an analog-to-digital (A/D) converter 154 for converting the analog signal from the microphone into a digital signal for processing, a DSP section 156 for filtering and compacting the digitized signal, a SCSI manager 158 for communication with the Macintosh, and a microphone control section 160 for controlling the microphone.

The Voice Navigator system also includes the Recognition Software voice drivers 120 which include routines for utterance detection 164 and command execution 166. For utterance detection 164, the voice drivers periodically poll 168 the Voice Navigator hardware to determine if an utterance is being received by Voice Navigator box 152, based on the amplitude of the signal received by the microphone. When an utterance is detected 170, the voice drivers create a speech buffer of encoded digital samples (tokens) to be used by the command execution drivers 166. On command 166 from the Voice Control driver 128, the recognition drivers can learn new utterances by token-to-terminal conversion 174. The token is converted to a macro for the utterance, and stored as a terminal in a voice file 122 (Fig. 1).

Recognition and pattern matching 172 is also performed on command by the voice drivers. During recognition, a stored token of incoming digitized samples is compared with macros for the utterances in the current level of the recognition hierarchy. If a match is found, terminal to output conversion 176 is also performed, selecting the command string associated with the

recognized utterance from the word list 124 (Fig. 1). State management 178, such as changing of sensitivity controls, is also performed on command by the voice drivers.

The Voice Control driver 128 forms an interface 182 to the voice drivers 120 through control commands, an interface 184 to the Macintosh operating system 132 (Fig. 1) through event posting and operating system hooks, and an interface 186 to the user through display menus and prompts.

The interface 182 to the drivers allows Voice Control access to the Voice Driver command functions 166. This interface allows Voice Control to monitor 188 the status of the recognizer, for example to check for an utterance token in the utterance queue buffered 170 to the Macintosh. If there is an utterance, and if processor time is available, Voice Control issues command `sdi_recognize` 190, calling the recognition and pattern match routine 172 in the voice drivers. In addition, the interface to the drivers may issue command `sdi_output` 192 which controls the terminal to output conversion routine 176 in the voice drivers, converting a recognized utterance to an command string for use by Voice Control. The command string may indicate mouse or keystroke events to be posted to the operating system, or may indicate commands to Voice Control itself (e.g. enabling or disabling Voice Control).

From the user's perspective, Voice Control is simply a Macintosh driver with internal parameters, such as sensitivity, and internal commands, such as commands to learn new utterances. The actual processing which the user perceives as Voice Control may actually be performed by Voice Control, or by the Voice Drivers, depending upon the function. For example, the utterance learning procedures are performed by the Voice Drivers under the control of Voice Control.

The interface 184 to the Macintosh operating system allows Voice Control, where appropriate, to manipulate the operating system (e.g., by posting events or modifying event queues). The macro interpreter 194 takes the command strings delivered from the

voice drivers via the text buffer and interprets them to decide what actions to take. These commands may indicate text strings to be displayed on the display or mouse movements or menu selections to be executed.

5 In the interpretive execution of the command strings, Voice Control must manipulate the Macintosh event queues. This task is performed by OS event management 196. As discussed above, voice events may simulate events which are ordinarily associated with the keyboard or with the mouse. Keyboard events are handled by OS event management 196 directly. Mouse events are handled by mouse handler 198. Mouse events require an additional level of handling because mouse events can require operating system manipulation outside of the standard event post routines which are accomplished by the OS event management 196.

10 The main interface into the Macintosh operating system 132 is event based, and is used in the majority of the commands which are voice recognized and issued to the Macintosh. However, there are other "hooks" to the operating system state which are used to control parameters such as mouse placement and mouse motion. For example, as will be discussed later, pushing the mouse button down generates an event, however, keeping the mouse button pushed down and dragging the mouse across a menu requires the use of an operating system hook. For reference, the operating system hooks used by the Voice Navigator are listed in Appendix B.

15 The operating system hooks are implemented by the trap filters 200, which are filters used by Voice Control to force the Macintosh operating system to accept the controls implemented by OS event management 196 and mouse handler 198.

20 The Macintosh operating system traps are held in Macintosh read only memories (ROMs), and implement high level commands for controlling the system. Examples of these high level commands are: drawing a string onto the screen, window zooming, moving windows to the front and back of the screen, and polling the status of the mouse button. In order for the Voice Control driver to properly interface with the Macintosh operating system it must control these

25

30

35

operating system traps to generate the appropriate events.

To generate menu events, for example, Voice Control "seizes" the menu select trap (i.e. takes control of the trap from the operating system). Once Voice Control has seized the trap, application requests for menu selections are forwarded to Voice Control. In this way Voice Control is able to modify, where necessary, the operating system output to the program, thereby controlling the system behavior as desired.

The interface 186 to the user provides user control of the Voice Control operations. Prompts 202 display the name of each recognized utterance on the Macintosh screen so that the user may determine if the proper utterance has been recognized. On-line training 204 allows the user to access, at any time while using the Macintosh, the utterance names in the word list 124 currently in use. The user may see which utterance names have been trained and may retrain the utterance names in an on-line manner (these functions require Voice Control to use the Voice Driver interface, as discussed above). User options 206 provide selection of various Voice Control settings, such as the sensitivity and confidence level of the recognizer (i.e., the level of certainty required to decide that an utterance has been recognized). The optimal values for these parameters depend upon the microphone in use and the speaking voice of the user.

The interface 186 to the user does not operate via the Macintosh event interface. Rather, it is simply a recursive loop which controls the Recognition Software and the state of the Voice Control driver.

Language Maker 140 includes an application analyzer 210 and an event recorder 212. Application analyzer 210 parses the executable code of applications as discussed above, and produces suitable default utterance names and pre-programmed command strings. The application analyzer 210 includes a menu extraction procedure 214 which searches executable code to find text strings corresponding to menus. The application analyzer 210 also includes control identification procedures 216 for creating the command

strings corresponding to each menu item in an application.

The event recorder 212 is a driver for recording user commands and creating command strings for utterances. This allows the user to easily create and edit command strings as discussed above.

Types of events which may be entered into the event recorder include: text entry 218, mouse events 220 (such as clicking at a specified place on the screen), special events 222 which may be necessary to control a particular application, and voice events 224 which may be associated with operations of the Voice Control driver.

Language Maker

Referring to Fig. 4, the Language Maker main event loop 230 is similar in structure to main event loops used by other desk accessories in the Macintosh operating system. If a desk accessory is selected from the "Apple" menu, an "open" event is transmitted to the accessory. In general, if the application in which it resides quits or if the user quits it using its menus, a "close" event is transmitted to the accessory. Otherwise, the accessory is transmitted control events. The message parameter of a control event indicates the kind of event. As seen in Fig. 4, the Language Maker main event loop 230 begins with an analysis 232 of the event type.

If the event is an open event Language Maker tests 234 whether it is already opened. If Language Maker is already opened 236, the current language (i.e. the list of utterance names from the current word list) is displayed and Language Maker returns 237 to the operating system. If Language Maker is not open 238, it is initialized and then returns 239 to the operating system.

If the event is a close event, Language Maker prompts the user 240 to save the current language as a language file. If the user commands Language Maker to save the current language, the current language is converted by the Write Production module 242 to a language file, and then Language Maker exits 244. If the current language is not saved, Language Maker exits directly.

If the event is a control event 246, then the way in which

Language Maker responds to the event depends upon the mode that Language Maker is in, because Language Maker has a utility for recording events (i.e. the mouse movements and clicks or text entry that the user wishes to assign to an utterance), and must record events which do not involve the Language Maker window. However, when not recording, Language Maker should only respond to events in its window. Therefore, Language Maker may respond to events in one mode but not in another.

A control event 246 is forwarded to one of three branches 248, 250, 252. All menu events are forwarded to the accMenu branch 252. (Only menu events occurring in desk accessory menus will be forwarded to Language Maker.) All window events for the Language Maker window are forwarded to the accEvent branch 250. All other events received by Language Maker, which correspond to events for desktop accessories or applications other than Language Maker, initiate activity in the accRun branch 248, to enable recording of actions.

In the accRun branch 248, events are recorded and associated with the selected utterance name. Before any events are recorded Language Maker checks 254 if Language Maker is recording; if not, Language Maker returns 256. If recording is on 258, then Language Maker checks the current recording mode.

While recording, Language Maker seizes control of the operating system by setting control flags that cause the operating system to call Language Maker every tick of the Macintosh (i.e. every 1/60 second).

If the user has set Language Maker in dialog mode, Language Maker can record dialog events (i.e. events which involve modal dialog, where the user cannot do anything except respond to the actions in modal dialog boxes). To accomplish this, the user must be able to produce actions (i.e. mouse clicks, menu selections) in the current application so that the dialog boxes are prompted to the screen. Then the user can initialize recording and respond to the dialog

boxes. When modal dialog boxes should be produced, events received by Language Maker are also forwarded to the operating system. Otherwise, events are not forwarded to the operating system. Language Maker's modal dialog recording is performed by the Run Modal module 260.

If modal dialog events are not being recorded, the user records with Language Maker in "action" mode, and Language Maker proceeds to the Run Edit module 262.

In the accEvent branch, all events are forwarded to the Event Handler module 264.

In the accMenu branch, the menu indicated by the desk accessory menu event is checked 266. If the event occurred in the Language Maker menu, it is forwarded to the Do My Menu module 268. Other events are ignored 270.

Referring to Fig. 5, the Run Edit module 262 performs a loop 272, 274. Each action is recorded by the Record Actions submodule 272. If there are more actions in the event queue then the loop returns to the Record Actions submodule. If a cancel action appears 276 in the event queue then Run Edit returns 277 without updating the current language in memory. Otherwise, if the events are completed successfully, run edit updates the language in memory and turns off recording 278 and returns to the operating system 280.

Referring to Fig. 6, in the Record Actions submodule 272, actions performed by the user in record mode are recorded. When the current application makes a request for the next event on the event queue, the event is checked by record actions. Each non-null event (i.e. each action) is processed by Record Actions. First, the type of action is checked 282. If the action selects a menu 284, then the selected menu is recorded. If the action is a mouse click 286, the In Button? routine (see Fig. 8) checks if the click occurred inside of a button (a button is a menu selection area in the front window) or not. If so, the button is recorded 288. If not, the location of the click is recorded 290.

Other actions are recorded by special handlers. These actions

include group actions 292, mouse down actions 294, mouse up actions 296, zoom actions 298, grow actions 300, and next window actions 302.

Some actions in menus can create pop-up menus with subchoices. These actions are handled by popping up the appropriate pop-up menu so that the user may select the desired subchoice. Move actions 304, pause actions 306, scroll actions 308, text actions 310 and voice actions 312 pop up respective menus and Record Actions checks 314 for the menu selection made by the user (with a mouse drag). If no menu selection is made, then no action is recorded 316. Otherwise, the choice is recorded 318.

Other actions may launch applications. In this case 320 the selected application is determined. If no application has been selected then no action is recorded 322, otherwise the selected application is recorded 324.

Referring to Fig. 7, the Run Modal procedure 260 allows recording of the modal dialogs of the Macintosh computer. During modal dialogs, the user cannot do anything except respond to the actions in the modal dialog box. In order to record responses to those actions, Run Modal has several phases, each phase corresponding to a step in the recording process.

In the first phase, when the user selects dialog recording, Run Modal prompts the user with a Language Maker dialog box that gives the user the options "record" and "cancel" (see Fig. 25). The user may then interact with the current application until arriving at the dialog click that is to be recorded. During this phase, all calls to Run Modal are routed through Select Dialog 326, which produces the initial Language Maker dialog box, and then returns 327, ignoring further actions.

To enter the second, recording, phase, the user clicks on the "record" button in the Language Maker dialog box, indicating that the following dialog responses are to be recorded. In this phase, calls to Run Modal are routed to Record 328, which uses the In Button? routine 330 to check if a button in current application's dialog box has been selected. If the click occurred in a button,

then the button is recorded 332, and Run Modal returns 333. Otherwise, the location of the click is recorded 334 and Run Modal returns 335.

Finally, when all clicks are recorded, the user clicks on the "cancel" button in the Language Maker dialog box, entering the third phase of the recording session. The click in the "cancel" button causes Run Modal to route to Cancel 336, which updates 338 the current language in memory, then returns 340.

Referring to Fig. 8, the In Button? procedure 286 determines whether a mouse click event occurred on a button. In Button? gets the current window control list 342 (a Macintosh global which contains the locations of all of the button rectangles in the current window, refer to Appendix B) from the operating system and parses the list with a loop 344 - 350. Each control is fetched 350, and then the rectangle of the control is found 346. Each rectangle is analyzed 348 to determine if the click occurred in the rectangle. If not, the next control is fetched 350, and the loop recurses. If, 344, the list is emptied, then the click did not occur on a button, and no is returned 352. However, if the click did occur in a rectangle, then, if, 351, the rectangle is named, the click occurred on a button, and yes is returned 354; if the rectangle is not named 356, the click did not occur on a button, and no is returned 356.

Referring to Fig. 9, the Event Handler module 264 deals with standard Macintosh events in the Language Maker display window. The Language Maker display window lists the utterance names in the current language. As shown in Fig. 9, Event Handler determines 358 whether the event is a mouse or keyboard event and subsequently performs the proper action on the Language Maker window.

Mouse events include: dragging the window 360, growing the window 362, scrolling the window 364, clicking on the window 368 (which selects an utterance name), and dragging on the window 370 (which moves an utterance name from one location on the screen to another, potentially changing the utterance's position in the language hierarchy). Double-clicking 366 on an utterance name in

the window selects that utterance name for action recording, and therefore starts the Run Edit module.

Keyboard events include the standard cut 372, copy 374, and paste 376 routines, as well as cursor movements down 380, up 382, right 384, and left 386. Pressing return at the keyboard 378, as with a double click at the mouse, selects the current utterance name for action recording by Run Edit. After the appropriate command handler is called, Event Handler returns 388. The modifications to the language hierarchy performed by the Event Handler module are reflected in hierarchical structure of the language file produced by the Write Production module during close and save operations.

Referring to Fig. 10, the Do My Menu module 268 controls all of the menu choices supported by Language Maker. After summoning the appropriate submodule (discussed in detail in Figs. 11A through 11I), Do My Menu returns 408.

Referring to Fig. 11A, the New submodule 390 creates a new language. The New submodule first checks 410 if Language Maker is open. If so, it prompts the user 412 to save the current language as a language file. If the user saves the current language, New calls Write Production module 414 to save the language. New then calls Create Global Words 416 and forms a new language 418. Create Global Words 416 will automatically enter a few global (i.e. resident in all languages) utterance names and command strings into the new language. These utterance names and command strings allow the user to make Voice Control commands, and correspond to utterances such as "show me the active words" and "bring up the voice options" (the utterance macros for the corresponding voice file are trained by the user, or copied from an existing voice file, after the new language is saved).

Referring to Fig. 11B, the Open submodule 392 opens an existing language for modification. The Open submodule 392 checks 420 if Language Maker is open. If so, it prompts the user 422 to save the current language, calling Write Production 424 if yes. Open then prompts the user to open the selected language 426. If

the user cancels, Open returns 428. Otherwise, the language is loaded 430 and Open returns 432.

Referring to Fig. 11C, the Save submodule 394 saves the current language in memory as a language file. Save prompts the user to save the current language 434. If the user cancels, Save returns 436, otherwise, Save calls Write Production 438 to convert the language into a state machine control file suitable for use by VOCAL (Fig. 2). Finally, Save returns 440.

Referring to Fig. 11D, the New Action submodule 396 initializes the event recorders to begin recording a new sequence of actions. New Action initializes the event recorder by displaying an action window to the user 442, setting up a tool palette for the user to use, and initializing recording of actions. Then New Action returns 444. After New Action is started, actions are not delivered to the operating system directly; rather they are filtered through Language Maker.

Referring to Fig. 11E, the Record Dialog submodule 398 records responses to dialog boxes through the use of the Run Modal module. Record Dialog 398 gives the user a way to record actions in modal dialog; otherwise the user would be prevented from performing the actions which bring up the dialog boxes. Record Dialog displays 446 the dialog action window (see Fig. 25) and turns recording on. Then Record Dialog returns 448.

Referring to Fig. 11F, the Create Default Menus submodule 400 extracts default utterance names (and generates associated command strings) from the executable code for an application. Create Default Menus 270 is ordinarily the first choice selected by a user when creating a language for a particular application. This submodule looks at the executable code of an application and creates an utterance name for each menu command in the application, associating the utterance name with a command string that will select that menu command. When called, Create Default Menus gets 450 the menu bar from the executable code of the application, and initializes the current menu to be the first menu (X=1). Next, each menu is processed recursively. When all menus are processed,

Create Default Menus returns 454. A first loop 452, 456, 458, 460 locates the current (X^{th}) menu handle 456, initializes menu parsing, checks if the current menu is fully parsed 458, and reiterates by updating the current menu to the next menu. A second loop 458, 462, 464 finds each menu name 462, and checks 464 if the name is hierarchical (i.e. if the name points to further menus). If the names are not hierarchical, the loop recurses. Otherwise, the hierarchical menu is fetched 466, and a third loop 470, 472 starts. In the third loop, each item name in the hierarchical menu is fetched 472, and the loop checks if all hierarchical item names have been fetched 470.

Referring to Fig. 11G, the Create Default Text submodule 402 allows the user to convert a text file on the clipboard into a list of utterance names. Create default text 402 creates an utterance name for each unique word in the clipboard 474, and then returns 476. The utterance names are associated with the keyboard entries which will type out the name. For example, a business letter can be copied from the clipboard into default text. Utterances would then be associated with each of the common business terms in the letter. After ten or twelve business letters have been converted the majority of the business letter words would be stored as a set of utterances.

Referring to Fig. 11H, the Alphabetize Group submodule 404 allows the user to alphabetize the utterance names in a language. The selected group of names (created by dragging the mouse over utterance names in the Language Maker window) is alphabetized 478, and then Alphabetize Group returns 480.

Referring to Fig. 11I, the Preferences submodule 406 allows the user to select standard graphic user interface preferences such as font style 482 and font size 484. The Preferences submenu 486 allows the user to state the metric by which mouse locations of recorded actions are stored. The coordinates for mouse actions can be relative to the global window coordinates or relative to the application window coordinates. In the case where application menu selections are performed by mouse clicks, the mouse clicks must

always be in relative coordinates so that the window may be moved on the screen without affecting the function of the mouse click. The Preferences submenu 486 also determines whether, when a mouse action is recorded, the mouse is left at the location of a click or returned to its original location after a click. When the preference selections are done 488, the user is prompted whether he wants to update the current preference settings for Language Maker. If so, the file is updated 490 and Preferences returns 492. If not, Preferences returns directly to the operating system 494 without saving.

Referring to Fig. 12, the Write Production module 242 is called when a file is saved. Write Production saves the current language and converts it from an outline processor format such as that used in the Language Maker application to a hierarchical text format suitable for use with the state machine based Recognition Software. Language files are associated with applications and new language files can be created or edited for each additional application to incorporate the various commands of the application into voice recognition.

The embodiment of the Write Production module depends upon the Recognition Software in use. In general, the Write Production module is written to convert the current language to suitable format for the Recognition Software in use. The particular embodiment of Write Production shown in Fig. 12 applies to the syntax of the VOCAL compiler for the Dragon Systems Recognition Software.

Write Production first tests the language 494 to determine if there are any sub-levels. If not, the Write Terminal submodule 496 saves the top level language, and Write Production returns 498. If sub-levels exist in the language, then each sub-level is processed by a tail-recursive loop. If a root entry exists in the language 500 (i.e. if only one utterance name exists at the current level) then Write Production writes 502 the string "Root = (" to the file, and checks for sub-levels 512. Otherwise, if no root exists, Write Terminal is called 504 to save the names in the

current level of the language. Next, the string "TERMINAL =" is written 506, and if, 508, the language level is terminal, the string "(" is written. Next, Write Production checks 512 for sub-levels in the language. If no sub-levels exist, Write Production returns 514. Otherwise, the sub-levels are processed by another call 516 to Write Production on the sub-level of the language. After the sub-level is processed, Write Production writes the string ")" and returns 518.

Referring to Fig. 13, the Write Terminal submodule 496 writes each utterance name and the associated command string to the language file. First, Write Terminal checks 520 if it is at a terminal. If not, it returns 530. Otherwise, Write Terminal writes 522 the string corresponding to the utterance name to the language file. Next, if, 524, there is an associated command string, Write Terminal writes the command string (i.e. "output") to the language file. Finally, Write Terminal writes 528 the string ";" to the language file and returns 530.

Voice Control

The Voice Control software serves as a gate between the operating system and the applications running on the operating system. This is accomplished by setting the Macintosh operating system's `get_next_event` procedure equal to a filter procedure created by Voice Control. The `get_next_event` procedure runs when each `next_event` request is generated by the operating system or by applications. Ordinarily the `get_next_event` procedure is null, and `next_event` requests go directly to the operating system. The filter procedure passes control to Voice Control on every request. This allows Voice Control to perform voice actions by intercepting mouse and keyboard events, and create new events corresponding to spoken commands.

The Voice Control filter procedure is shown in Fig. 14.

After installation 538, the `get_next_event` filter procedure 540 is called before an event is generated by the operating system. The event is first checked 542 to see if it is a null event. If so, the Process Input module 544 is called directly. The Process

Input routine 544 checks for new speech input and processes any that has been received. After Process Input, the Voice Control driver proceeds through normal filter processing 546 (i.e., any filter processing caused by other applications) and returns 548.

5 If the next event is not a null event, then displays are hidden 550. This allows Voice Control to hide any Voice Control displays (such as current language lists) which could have been generated by a previous non-null action. Therefore, if any prompt windows have been produced by Voice Control, when a non-null event occurs, 10 the prompt windows are hidden. Next, key down events are checked 552. Because the recognizer is controlled (i.e. turned on and off) by certain special key down events, if the event is a key down event then Voice Control must do further processing. Otherwise, the Voice Control drive procedure moves directly to Process Input 544. 15 If a key down event has occurred 554, where appropriate, software latches which control the recognizer are set. This allows activation of the Recognizer Software, the selection of Recognizer options, or the display of languages. Thereafter, the Voice Control driver moves to Process Input 544.

20 Referring to Fig. 15, the Process Input routine is the heart of the Voice Control driver. It manages all voice input for the Voice Navigator. The Process Input module is called each time an event is processed by the operating system. First 546, any latches which need to be set are processed, and the Macintosh waits for a number of delay ticks, if necessary. Delay ticks are included, for 25 example, where a menu drag is being performed by Voice Control, to allow the menu to be drawn on the screen before starting the drag. Also, some applications require delay between mouse or keyboard events. Next, if recognition is activated 548 the process input routine proceeds to do recognition 562. 30 If recognition is deactivated, Process Input returns 560.

The recognition routine 562 prompts the recognition drivers to check for an utterance (i.e., sound that could be speech input). If there is recognized speech input 564, Process Input checks the 35 vertical blanking interrupt VBL handler 566, and deactivates it

where appropriate.

The vertical blanking interrupt cycle is a very low level cycle in the operating system. Every time the screen is refreshed, as the raster is moving from the bottom right to the top left of the screen, the vertical blanking interrupt time occurs. During this blanking time, very short and very high priority routines can be executed. The cycle is used by the Process Input routine to move the mouse continuously by very slowly incrementing of the mouse coordinates where appropriate. To accomplish this, mouse move events are installed onto the VBL queue. Therefore, where appropriate, the VBL handler must be deactivated to move the mouse.

Other speech input is placed 568 on a speech queue, which stores speech related events for the processor until they can be handled by the ProcessQ routine. However, regardless of whether speech is recognized, ProcessQ 570 is always called by Process Input. Therefore, the speech events queued to ProcessQ are eventually executed, but not necessarily in the same Process Input cycle. After calling ProcessQ, Process Input returns 571.

Referring to Fig. 16, the Recognize submodule 562 checks for encoded utterances queued by the Voice Navigator box, and then calls the recognition drivers to attempt to recognize any utterances. Recognize returns the number of commands in (i.e. the length of) the command string returned from the recognizer. If, 572, no utterance is returned from the recognizer, then Recognize returns a length of zero (574), indicating no recognition has occurred. If an utterance is available, then Recognize calls sdi_recognize 576, instructing the Recognizer Software to attempt recognition on the utterance. If, 578, recognition is successful, then the name of the utterance is displayed 582 to the user. At the same time, any close call windows (i.e. windows associated with close call choices, prompted by Voice Control in response to the Recognizer Software) are cleared from the display. If recognition is unsuccessful, the Macintosh beeps 580 and zero length is returned 574.

If recognition is successful, Recognize searches 584 for an

output string associated with the utterance. If there is an output string, recognize checks if it is asleep 586. If it is not asleep 590, the output count is set to the length of the output string and, if the command is a control command 592 (such as "go to sleep" or "wake up"), it is handled by the Process Voice Commands routine 594.

If there is no output string for the recognized utterance, or if the recognizer is asleep, then the output of Recognize is zero (588). After the output count is determined 596, the state of the recognizer is processed 596. At this time, if the Voice Control state flags have been modified by any of the Recognize subroutines, the appropriate actions are initialized. Finally, Recognize returns 598.

Referring to Fig. 17, the Process Voice Commands module deals with commands that control the recognizer. The module may perform actions, or may flag actions to be performed by the Process States block 596 (Fig. 16). If the recognizer is put to sleep 600 or awakened 604, the appropriate flags are set 602, 606, and zero is returned 626, 628 for the length of the command string, indicating to Process States to take no further actions. Otherwise, if the command is scratch_that 608 (ignore last utterance), first_level 612 (go to top of language hierarchy, i.e. set the Voice Control state to the root state for the language), word_list 616 (show the current language), or voice_options 620, the appropriate flags are set and 610, 614, 618, 622, and a string length of -1 is returned 624, 628, indicating that the recognizer state should be changed by Process States 596 (Fig. 16).

Referring to Fig. 18 the ProcessQ module 570 pulls speech input from the speech queue and processes it. If, 630, the event queue is empty then ProcessQ may proceed, otherwise ProcessQ aborts 632 because the event queue may overflow if speech events are placed on the queue along with other events. If, 634, the speech queue has any events then process queue checks to see if, 636, delay ticks for menu drawing or other related activities have expired. If no events are on the speech queue the ProcessQ aborts

636. If delay ticks have expired, then ProcessQ calls Get Next 642 and returns 644. Otherwise, if delay ticks have not expired, ProcessQ aborts 640.

Referring to Fig. 19, the Get Next submodule 642 gets
 5 characters from the speech queue and processes them. If, 646, there are no characters in the speech queue then the procedure simply returns 648. If there are characters in the speech queue then Get Next checks 650 to see if the characters are command characters. If they are, then Get Next calls Check Command 660.
 10 If not, then the characters are text, and Get Next sets the meta bits 652 where appropriate.

When the Macintosh posts an event, the meta bits (see Appendix B) are used as flags for conditioning keystrokes such as the condition key, the option key, or the command key. These keys condition the character pressed at the keyboard and create control characters. To create the proper operating system events, therefore, the meta bits must be set where necessary. Once
 15 the meta bits are set 652, a key down event is posted 654 to the Macintosh event queue, simulating a keypush at the keyboard. Following this, a key up is posted 656 to the event queue, simulating a key up. If, 658, there is still room in the event queue, then further speech characters are obtained and processed 646. If not, then the Get Next procedure returns 676.

If the command string input corresponds to a command rather
 25 than simple key strokes, the string is handled by the Check Command procedure 660 as illustrated in Figure 19. In the Check Command procedure 660 the next four characters from the speech queue (four characters is the length of all command strings, see Appendix A) are fetched 662 and compared 664 to a command table. If, 666, the
 30 characters equal a voice command, then a command is recognized, and processing is continued by the Handle Command routine 668. Otherwise, the characters are interpreted as text and processing returns to the meta bits step 652.

In the Handle Command procedure 668 each command is referenced
 35 into a table of command procedures by first computing 670 the

command handler offset into the table and then referencing the table, and calling the appropriate command handler 672. After calling the appropriate command handler, Get Next exits the Process Input module directly 674 (the structure of the software is such that a return from Handle Command would return to the meta bits step 652, which would be incorrect).

The command handlers available to the Handle Command routine are illustrated in Figure 20. Each command handler is detailed by a flow diagram in Figures 21A through 21G. The syntax for the commands is detailed in Appendix A.

Referring to Figure 21A, the Menu command will pull down a menu, for example, @MENU(apple,0) (where apple is the menu number for the apple menu) will pull down the apple menu. Menu command will also select an item from the menu, for example, @MENU(apple,calculator) (where calculator is the itemnumber for the calculator in the apple menu) will select the calculator from the apple menu. Menu command initializes by running the Find Menu routine 678 which queues the menu id and the item number for the selected menu. (If the item number in the menu is 0 then Find Menu simply clicks on the menu bar.) After Find Menu returns, if 680, there are no menus queued for posting, the Menu command simply returns 690. However, if menus are queued for posting, Menu command intercepts 682 one of the Macintosh internal traps called Menu Select. The Menu Select trap is set equal to the My Menu Select routine 692. Next the cursor coordinates are hidden 684 so that the mouse cannot be seen as it moves on the screen. Next, Menu command posts 686 a mouse down (i.e. pushes the mouse button down) on the menu bar. When the mouse down occurs on the menu bar the Macintosh operating system generates a menu event for the application. Each application receiving a menu event requests service from the operating system to find out what the menu event is. To do this the application issues a Menu Select trap. The menu select trap then places the location of the mouse on the stack. However, when the application issues a menu select trap in this case, it is serviced by the My Menu Select routine 692

instead, thereby allowing Menu command to insert the desired menu coordinates in the place of the real coordinates. After posting a mouse down in the appropriate menu bar, Menu Command sets 688 the wait ticks to 30, which gives the operating system time to draw the menu, and returns 690.

In the My Menu Select trap 692 the menuselect global state is reset 694 to clear any previously selected menus, and the desired menu id and the item number are moved to the Macintosh stack 696, thus selecting the desired menu item.

The Find Menu routine 700 collects 702 the command parameters for the desired menu. Next, the menuname is compared 704 to the menu name list. If, 706, there is no menu with the name "menuname", Find Menu exits 708. Otherwise, Find Menu compares 710 the itemname to the names of the items in the menu. If, 712, the located item number is greater than 0, then Find Menu queues 718 the menu id and item number for use by Menu command, and returns 720. Otherwise, if the item number is 0 then Find Menu simply sets 714 the internal Voice Control flags "mousedown" and "global" flags to true. This indicates to Voice Control that the mouse location should be globally referenced, and that the mouse button should be held down. Then Find Menu calls 716 the Post Mouse routine, which references these flags to manipulate the operating system's mouse state accordingly.

Referring to Fig. 21B, the Control command 722 performs a button push within a menu, invoking actions such as the save command in the file menu of an application. To do this, the Control command gets the command parameters 724 from the control string, finds the front window 726, gets the window command list 728, and checks 730 if the control name exists in the control list. If the control name does exist in the control list then the control rectangle coordinates are calculated 732, the Post Mouse routine 734 clicks the mouse in the proper coordinates, and the Control command returns 736. If the control name is not found, the Control command returns directly.

The Keypad command 738 simulates numerical entries at the

Macintosh keypad. Keypad finds the command parameters for the command string 740, gets the keycode value 742 for the desired key, posts a key down event 744 to the Macintosh event queue, and returns 746.

5 The Zoom command 748 zooms the front window. Zoom obtains the front window pointer 750 in order to reference the mouse to the front window, calculates the location of the zoom box 752, uses Post Mouse to click in the zoom box 754, and returns 756.

10 The Local Mouse command 758 clicks the mouse at a locally referenced location. Local Mouse obtains the command parameters for the desired mouse location 760, uses Post Mouse to click at the desired coordinate 762, and returns 764.

15 The Global Mouse command 766 clicks the mouse at a globally referenced location. Global Mouse obtains the command parameters for the desired mouse location 768, sets the global flag to true 770 (to signal to Post Mouse that the coordinates are global), uses Post Mouse to click at the desired coordinate 772, and returns 774.

20 The Double Click command double clicks the mouse at a locally referenced location. Double Click obtains the command parameters for the desired mouse location 778, calls Post Mouse twice 780, 782 (to click twice in the desired location), and returns 784.

25 The Mouse Down command 786 sets the mouse button down. Mouse Down sets the mousedown flag to true 788 (to signal to Post Mouse that mouse button should be held down), uses Post Mouse to set the button down 790, and returns 792.

30 The Mouse Up command 794 sets the mouse button up. Mouse Up sets the mbState global (see Appendix B) to Mouse Button UP 796 (to signal to the operating system that mouse button should be set up), posts a mouse up event to the Macintosh event queue 798 (to signal to applications that the mouse button has gone up), and returns 800.

35 Referring to Fig. 21D, the Screen Down command 802 scrolls the contents of the current window down. Screen Down first looks 804 for the vertical scroll bar in the front window. If, 806, the scroll bar is not found, Screen Down simply returns 814. If the

scroll bar is found, Screen Down calculates the coordinates of the down arrow 808, sets the mousedown flag to true 810 (indicating to Post Mouse that the mouse button should be held down), uses Post Mouse to set the mouse button down 812, and returns 814.

5 The Screen Up command 816 scrolls the contents of the current window up. Screen Up first looks 818 for the vertical scroll bar in the front window. If, 820, the scroll bar is not found, Screen Up simply returns 828. If the scroll bar is found, Screen Up calculates the coordinates of the up arrow 822, sets the mousedown flag to true 824 (indicating to Post Mouse that the mouse button should be held down), uses Post Mouse to set the mouse button down 826, and returns 828.

15 The Screen Left command 830 scrolls the contents of the current window left. Screen Left first looks 832 for the horizontal scroll bar in the front window. If, 834, the scroll bar is not found, Screen Left simply returns 842. If the scroll bar is found, Screen Left calculates the coordinates of the left arrow 836, sets the mousedown flag to true 838 (indicating to Post Mouse that the mouse button should be held down), uses Post Mouse to set the mouse button down 840, and returns 842.

20 The Screen Right command 844 scrolls the contents of the current window right. Screen Right first looks 846 for the horizontal scroll bar in the front window. If, 848, the scroll bar is not found, Screen Right simply returns 856. If the scroll bar is found, Screen Right calculates the coordinates of the right arrow 850, sets the mousedown flag to true 852 (indicating to Post Mouse that the mouse button should be set down), uses Post Mouse to set the mouse button down 854, and returns 856.

25 Referring to Fig. 21E, the Page Down command 858 moves the contents of the current window down a page. Page Down first looks 30 860 for the vertical scroll bar in the front window. If, 862, the scroll bar is not found, Page Down simply returns 868. If the scroll bar is found, Page Down calculates the page down button coordinates 864, uses Post Mouse to click the mouse button down 35 866, and returns 868.

The Page Up command 870 moves the contents of the current window up a page. Page Up first looks 872 for the vertical scroll bar in the front window. If, 874, the scroll bar is not found, Page Up simply returns 880. If the scroll bar is found, Page Up
 5 calculates the page up button coordinates 876, uses Post Mouse to click the mouse button down 878, and returns 880.

The Page Left command 882 moves the contents of the current window left a page. Page Left first looks 884 for the horizontal scroll bar in the front window. If, 886, the scroll bar is not
 10 found, Page Left simply returns 892. If the scroll bar is found, Page Left calculates the page left button coordinates 888, uses Post Mouse to click the mouse button down 890, and returns 892.

The Page Right command 894 moves the contents of the current window right a page. Page Right first looks 896 for the horizontal scroll bar in the front window. If, 898, the scroll bar is not
 15 found, Page Right simply returns 904. If the scroll bar is found, Page Right calculates the page right button coordinates 900, uses Post Mouse to click the mouse button down 902, and returns 904.

Referring to Fig. 21F, the Move command 906 moves the mouse from its current location (y,x), to a new location (y+ δy , x+ δx). First, Move gets the command parameters 908, then Move sets the mouse speed to tablet 910 (this cancels the mouse acceleration,
 20 which otherwise would make mouse movements uncontrollable), adds the offset parameters to the current mouse location 912, forces a new cursor position and resets the mouse speed 914, and returns 916.

The Move to Global Coordinate command 918 moves the cursor to the global coordinates given by the Voice Control command string. First, Move to Global gets the command parameters 920, then Move
 30 to Global checks 922 if there is a position parameter. If there is a position parameter, the screen position coordinates are fetched 924. In either case, the global coordinates are calculated 926, the mouse speed is set to tablet 928, the mouse position is set to the new coordinates 930, the cursor is forced to the new
 35 position 932, and Move to Global returns 934.

The Move to Local Coordinate command 936 moves the cursor to the local coordinates given by the Voice Control command string. First, Move to Local gets the command parameters 938, then Move to Local checks 940 if there is a position parameter. If there is a position parameter, the local position coordinates are fetched 942. In either case, the global coordinates are calculated 944, the mouse speed is set to tablet 946, the mouse position is set to the new coordinates 948, the cursor is forced to the new position 950, and Move to Global returns 952.

The Move Continuous command 954 moves the mouse continuously from its present location, moving $\delta y, \delta x$ every refresh of the screen. This is accomplished by inserting 956 the VBL Move routine 960 in the Vertical Blanking Interrupt queue of the Macintosh and returning 958. Once in the queue, the VBL Move routine 960 will be executed every screen refresh. The VBL Move routine simply adds the δy and δx values to the current cursor position 962, resets the cursor 964, and returns 966.

Referring to Fig. 21G, the Option Key Down command 968 sets the option key down. This is done by setting the option key bit in the keyboard bit map to TRUE 970, and returning 972.

The Option Key Up command 974 sets the option key up. This is done by setting the option key bit in the keyboard bit map to FALSE 976, and returning 978.

The Shift Key Down command 980 sets the shift key down. This is done by setting the shift key bit in the keyboard bit map to TRUE 982, and returning 984.

The Shift Key Up command 986 sets the shift key up. This is done by setting the shift key bit in the keyboard bit map to FALSE 988, and returning 990.

The Command Key Down command 992 sets the command key down. This is done by setting the command key bit in the keyboard bit map to TRUE 994, and returning 996.

The Command Key Up command 998 sets the command key up. This is done by setting the command key bit in the keyboard bit map to FALSE 1000, and returning 1002.

The Control Key Down command 1004 sets the control key down. This is done by setting the control key bit in the keyboard bit map to TRUE 1006, and returning 1008.

5 The Control Key Up command 1010 sets the control key up. This is done by setting the control key bit in the keyboard bit map to FALSE 1012, and returning 1014.

The Next Window command 1016 moves the front window to the back. This is done by getting the front window 1018 and sending it to the back 1020, and returning 1022.

10 The Erase command 1024 erases numchars characters from the screen. The number of characters typed by the most recent voice command is stored by Voice Control. Therefore, Erase will erase the characters from the most recent voice command. This is done by a loop which posts delete key keydown events 1026 and checks 15 1028 if the number posted equals numchars. When numchars deletes have been posted, Erase returns 1030.

The Capitalize command 1032 capitalizes the next keystroke. This is done by setting the caps flag to TRUE 1034, and returning 1036.

20 The Launch command 1038 launches an application. The application must be on the boot drive no more than one level deep. This is done by getting the name of the application 1040 ("appl_name"), searching for appl_name on the boot volume 1042, and, if, 1044, the application is found, setting the volume to the application folder 1048, launching the application 1050 (no return 25 is necessary because the new application will clear the Macintosh queue). If the application is not found, Launch simply returns 1046.

30 Referring to Fig. 22, the Post Mouse routine 1052 posts mouse down events to the Macintosh event queue and can set traps to monitor mouse activity and to keep the mouse down. The actions of Post Mouse are determined by the Voice Control flags global and mousedown, which are set by command handlers before calling Post Mouse. After a Post Mouse, when an application does a 35 get_next_event it will see a mouse down event in the event queue,

leading to events such as clicks, mouse downs or double clicks.

First, Post Mouse saves the current mouse location 1054 so that the mouse may be returned to its initial location after the mouse events are produced. Next the cursor is hidden 1056 to shield the user from seeing the mouse moving around the screen. Next the global flag is checked. If, 1058, the coordinates are local (i.e. global = FALSE) then they are converted 1060 to global coordinates. Next, the mouse speed is set to tablet 1062 (to avoid acceleration problems), and the mouse down is posted to the Macintosh event queue 1064. If, 1066, the mousedown flag is TRUE (i.e. if the mouse button should be held down) then the Set Mouse Down routine is called 1072 and Post Mouse returns 1070. Otherwise, if the mouse down flag is FALSE, then a click is created by posting a mouse up event to the Macintosh event queue 1068 and returning 1070.

Referring to Fig. 23, the Set Mouse Down routine 1072 holds the mouse button down by replacing 1074 the Macintosh button trap with a Voice Control trap named My Button. The My Button trap then recognizes further voice commands and creates mouse drags or clicks as appropriate. After initializing My Button, Set Mouse Down checks 1076 if the Macintosh is a Macintosh Plus, in which case the Post Event trap must also be reset 1078 to the Voice Control My Post Event trap. (The Macintosh Plus will not simply check the mbState global flag to determine the mouse button state. Rather, the Post Event trap in a Macintosh Plus will poll the actual mouse button to determine its state, and will post mouse up events if the mouse button is up. Therefore, to force the Macintosh Plus to accept the mouse button state as dictated by Voice Control, during voice actions, the Post Event trap is replaced with a My Post Event trap, which will not poll the status of the mouse button.) Next, the mbState flag is set to MouseDown 1080 (indicating that the mouse button is down) and Set Mouse Down returns 1082.

The My Button trap 1084 replaces the Macintosh button trap, thereby seizing control of the button state from the operating system. Each time My Button is called, it checks 1086 the

Macintosh mouse button state bit mbState. If mbState has been set to UP, My Button moves to the End Button routine 1106 which sets mbState to UP 1108, removes any VBL routine which has been installed 1110, resets the Button and Post Event traps to the original Macintosh traps 1112, resets the mouse speed and couples the cursor to the mouse 1114, shows the cursor 1102, and returns 1104.

However, if the mouse button is to remain down, My Button checks for the expiration of wait ticks (which allow the Macintosh time to draw menus on the screen) 1088, and calls the recognize routine 1090 to recognize further speech commands. After further speech commands are recognized, My Button determines 1092 its next action based on the length of the command string. If the command string length is less than zero, then the next voice command was a Voice Control internal command, and the mouse button is released by calling End Button 1106. If the command string length is greater than zero, then a command was recognized, and the command is queued onto the voice que 1094, and the voice queue is checked for further commands 1096. If nothing was recognized (command string length of zero), then My Button skips directly to checking the voice queue 1096. If there is nothing in the voice queue, then My Button returns 1104. However, if there is a command in the voice queue, then My Button checks 1098 if the command is a mouse movement command (which would cause a mouse drag). If it is not a mouse movement, then the mouse button is released by calling End Button 1106. If the command is a mouse movement, then the command is executed 1100 (which drags the mouse), the cursor is displayed 1102, and My Button returns.

Screen Displays

Referring to Fig. 24, a screen display of a record actions session is shown. The user is recording a local mouse click 1106, and the click is being acknowledged in the action list 1108 and in the action window 1110.

Referring to Fig. 25, a record actions session using dialog boxes is shown. The dialog boxes 1112 for recording a manual

printer feed are displayed to the user, as well as the Voice Control Run Modal dialog box 1114 prompting the user to record the dialogs. The user is preparing to record a click on the Manual Feed button 1116.

Referring to Fig. 26, the Language Maker menu 1118 is shown.

Referring to Fig. 27, the user has requested the current language, which is displayed by Voice Control in a pop-up display 1120.

Referring to Fig. 28, the user has clicked on the utterance name "apple" 1122, requesting a retraining of the utterance for "apple". Voice Control has responded with a dialog box 1124 asking the user to say "apple" twice into the microphone.

Referring to Fig. 29, the text format of a Write Production output file 1126 (to be compiled by VOCAL) and the corresponding Language Maker display for the file 1128 are shown. It is clear from Fig. 29 that the Language Maker display is far more intuitive.

Referring to Fig. 30, a listing of the Write Production output file as displayed in Fig. 29 is provided.

Other Embodiments

Other embodiments of the invention are within the scope of the claims which follow the appendices. For example, the graphic user interface controlled by a voice recognition system could be other than that of the Apple Macintosh computer. The recognizer could be other than that marketed by Dragon Systems.

Included in the Appendices are Appendix A, which sets forth the Voice Control command language syntax, Appendix B, which lists some of the Macintosh OS globals used by the Voice Navigator system, Appendix C, which is a fiche of the Voice Navigator executable code, Appendix D, which is the Developer's Reference Manual for the Voice Navigator system, and Appendix E, which is the Voice Navigator User's Manual, all incorporated by reference herein.

A portion of the disclosure of this patent document contains material which is subject to copyright protection (for example, the microfiche Appendix, the User's Manual, and the Reference

[illegible]

Appendix A: Voice Control Command Language Syntax

Menu Command - @MENU(menuname,itemnum).

Finds item named itemnum in the menu named menuname and selects it. If itemnum is 0, hold the menu down.

5 Control Command - @CTRL(ctlname)

Finds the control named ctlname and clicks in its rectangle.

Key Pad Command - @KYPD(n), where n = 0-9, -, +, *, /, =, and
c for clear

Posts a Keydown for keys on the numeric keypad.

10 Zoom Command - @ZOOM

Clicks in the zoom box of the front window.

Local Mouse Click Command - @LMSE(y,x)

Clicks at local coordinates (y,x) of the front window.

Global Mouse Click Command - @GMSE(y,x)

15 Clicks at the global coordinates (y,x) of the current screen.

Double Click Command - @DCLK(y,x)

Double clicks at the global coordinates (y,x) of the current screen. If y=x=0, double click at the current Mouse location.

Mouse Down Command - @MSDN

20 Set the mouse button state to down and set up traps to keep
it down.

Mouse Up Command - @MSUP

Set the mouse button state to up.

Scroll Down Command - @SCDN

25 Post a mouse down in the down arrow portion of the front

window's scroll bar.

Scroll Up Command - @SCUP

Post a mouse down in the up arrow portion of the front window's scroll bar.

5 Scroll Left Command - @SCUP

Post a mouse down in the left arrow portion of the front window's scroll bar.

Scroll Right Command - @SCRT

10 Post a mouse down in the right arrow portion of the front window's scroll bar.

Page Down Command - @PGDN

Click in the page down portion of the front window's scroll bar.

15 Page Up Command - @PGUP

Click in the page up portion of the front window's scroll bar.

Page Left Command - @PGLF

Click in the page left portion of the front window's scroll bar.

20 Page Right Command - @PGRT

Click in the page right portion of the front window's scroll bar.

Move Command - @MOVE(δy , δx)

25 Move the Mouse from its current location (y, x), to a new location ($y+\delta y, x+\delta x$) where δy and δx are pixels and can be either positive or negative values.

Move Continuous Command - MOVI(δy , δx)

Move the mouse continuously from its present location, moving $\delta y, \delta x$ every refresh of the screen.

Move to Local Coordinate Command - `MOVL(y,x<,windowname>)` or `MOVL(n<,y,x<,windowname>>` where $n=N,S,E,W,NE,SE,SW,NW,C,G$

5 Move the cursor to the local coordinates given by (y,x) or by $(n.v+y,n.h+x)$. Use the `grafPort` of the window named "windowname".
If there is no "windowname" use the `grafPort` of the front window.

Move to Global Coordinate Command - `@MOVG(n,<y,x>)`

where $n=N,S,E,W,NE,SE,SW,NW,C,G$

10 move the cursor to the global coordinates given by (y,x) or by $(n.v+y,n.h+x)$. Use the `grafPort` of the screen.

Option Key Down Command - `@OPTD`

Press (and hold) the option key.

Option Key Up Command - `@OPTU`

15 Release the option key.

Shift Key Down Command - `@SHFD`

Press (and hold) the shift key.

Shift Key Up Command - `@SHFU`

Release the shift key.

20 Command Key Down Command - `@CMDD`

Press (and hold) the command key.

Command Key Up Command - `@CMDU`

Release the command key.

Control Key Down Command - `@CTLD`

25 Press (and hold) the control key.

5

9

IO

[illegible]

1. 100
 2. 200
 3. 300
 4. 400
 5. 500
 6. 600
 7. 700
 8. 800
 9. 900
 10. 1000

1111

100

10

Appendix B: Macintosh OS Globals

Interfacing to the Macintosh Operating System requires that certain low memory globals be managed by Voice Control. The following describes the most important globals. Further information is available in "Inside Macintosh", Vols. I-V.

Mouse Globals

MickeyBytes EQU \$D6A - a pointer to the cursor value; used to control the acceleration of the mouse. Set to point to tablet whenever the mouse is moved more than 10 pixels. [pointer]

MTEmp EQU \$828 - a low-level interrupt mouse location; used to move the mouse during VBL handling while executing a @MOVI command. [long]

Mouse EQU \$830 - the processed mouse coordinate; used to move the mouse for all other @MOVx commands. [long]

MBState EQU \$172 - current mouse button state; used to set the MouseDown for @MSDN and for @MENU when itemname = 0. [byte]

Keyboard Globals

KeyMap EQU \$174 - keyboard bit map, with one bit mapped to each key on the keyboard. Set the bit to TRUE to set the Meta keys (option, command, shift, control) down. [2 longs]

Filter Globals

JGNEFilter EQU \$29A - Get Next Event filter proc; set to Voice Control's main loop to intercept calls to Get Next Event. [pointer]

Event Queue Globals

evtMax EQU \$1E - maximum number of events in the event queue. When this number is reached, stop Posting events.

EventQueue EQU \$14A - event queue header, the location of the Macintosh event queue. [10 bytes]

Time Globals

5 Ticks EQU \$16A - Tick count, time since boot. Used to measure elapsed time between Voice Control actions. [long]

Cursor Globals

CrshrCouple EQU \$8CF - cursor coupled to mouse? Used to disconnect cursor when doing remote clicks with @LMSE and @GMSE. [byte]

10 CrshrNew EQU \$8CE - Cursor changed? Force a new cursor after moving the cursor. [byte]

Menu Globals

15 MenuList EQU \$A1 Current menuBar list structure. This handle can be de-referenced to find all the menus associated with an application. Use for @MENU commands [handle]

Window Globals

20 WindowList EQU \$9D6 - Z-ordered linked list of windows. This pointer will lead to a chain of all existing windows for an application. Use to find a window queue for all local commands. [pointer]

Window Offsets

These values are offsets within the window records that describe characteristics of the window. Once a window is located, these offsets are used to calculate:

25 thePort EQU 0 - GrafPtr; local coordinates for @LMSE and @MOVL commands.

portRect EQU \$10 - port's rectangle [rect]; window relative forms of the @MOVL command.

controlList EQU 140 - used to find the controls associated

5

5

5

5